Package: rankFD (via r-universe)

September 9, 2024

Type Package

Title Rank-Based Tests for General Factorial Designs

Version 0.1.1

Date 2022-03-23

Author Frank Konietschke, Sarah Friedrich, Edgar Brunner, Markus Pauly

Maintainer Frank Konietschke <frank.konietschke@charite.de>

Depends R (>= 3.2.2)

Description The rankFD() function calculates the Wald-type statistic (WTS) and the ANOVA-type statistic (ATS) for nonparametric factorial designs, e.g., for count, ordinal or score data in a crossed design with an arbitrary number of factors. Brunner, E., Bathke, A. and Konietschke, F. (2018) <doi:10.1007/978-3-030-02914-2>.

License GPL-2 | GPL-3

Imports lattice (>= 0.20-33), MASS (>= 7.3-43), coin (>= 1.1-2), multcomp, mvtnorm

LazyData TRUE

RoxygenNote 7.1.2

Encoding UTF-8

NeedsCompilation no

Repository CRAN

Date/Publication 2022-03-23 10:10:08 UTC

Contents

Coal	*Coal Acidity*

Description

Coal acidity values determined under each of three NaOH concentration levels for two different samples from each type of coal

Usage

```
data(Coal)
```

Format

A data frame with 18 rows and 3 variables:

Acidity resulting acidity values

NaOH the NaOH concentration

Type three different types of coal: "Morwell", "Yallourn" and "Maddingley"

Source

Hollander, M., Wolfe, D. A., Chicken, E. (2014) *Nonparametric Statistical Methods*. Wiley Series in Probability and Statistics.

Sternhell, S. (1958) Chemistry of brown coals VI: Further aspects of the chemistry of hydroxyl groups in Victorian brown coals. Australian Journal of Applied Science **9**, 375–379.

Muco	*Half-Time of Mucociliary Clearance*

Description

Mucociliary efficiency was assessed from the rate of removal of dust in three different groups of subjects

Usage

```
data(Muco)
```

Format

A data frame with 14 rows and 2 variables:

HalfTime Half-Time of Mucociliary clearance, assessed from the rate of removal of dust

Disease normal subjects, subjects with obstructive airways disease (OAD) and subjects with asbestosis

Source

Hollander, M., Wolfe, D. A., Chicken, E. (2014) Nonparametric Statistical Methods. Wiley Series in Probability and Statistics.

Thomson, M. L. and Short, M. D.(1969) Mucociliary function in health, chronic obstructive airway disease, and asbestosis. Journal of Applied Physiology **26**, 535–539.

---

nms   *Irritation of Nasal Mucosa*

Description

Damage of two gaseous substances on nasal mucosa membrane of mice

Usage

```
data(nms)
```

Format

A data frame with 150 rows and 3 variables:

subst The substance given, either 1 or 2

conc the concentration in which the substance was given, 1, 2 or 5 ppm

score degree of irritation assessed using an ordinal score ranging from 0 to 4 with 0 = "no irritation", 1 = "mild irritation", 2 = "strong irritation", 3 = "severe irritation" and 4 = "irreversible damage"

Source

Brunner, E., Bathke, A.C., Konietschke, F. Rank and Pseudo-Rank Procedures for Independent Observations in Factorial Designs. Springer International Publishing, 2018.

noether	*Sample size calculation for the Wilcoxon-Mann-Whitney test using the Noether formula. The function estimates the sample size needed to detect the effect with pre-defined power at significance level alpha using Noether's formula'.*

Description

Sample size calculation for the Wilcoxon-Mann-Whitney test using the Noether formula. The function estimates the sample size needed to detect the effect with pre-defined power at significance level alpha using Noether's formula'.

Usage

```
noether(alpha, power, t, p, x1 = c(0), ties = FALSE)
```

Arguments

alpha	two sided type I error rate
power	power: detect a relative effect p at least with probability power
t	proportion of subjects in the first group (between 0 and 1)
p	relative effect
x1	advance information is only needed in case of ties
ties	TRUE if ties are possible (non continuous distribution), otherwise FALSE

Value

Returns a data frame with the sample sizes for each group

References

Noether, G. E. (1987). Sample Size Determination for Some Common Nonparametric Tests. Journal of the American Statistical Association 85, 645.647.

Examples

```
noether(0.05,0.8,1/2, 0.75)
```

| psr | *A function for computing pseudo-ranks of data* |

Description

The `psr()` function calculates pseuo-ranks of data in general factorial designs. It returns the input data set complemented by an additional variable containing the pseudo-ranks. We note that more efficient algorithms for the computation of pseudo-ranks are implemented within the package pseudorank.

Usage

```
psr(formula, data, psranks = "pseudorank")
```

Arguments

formula	A model [formula](#) object. The left hand side contains the response variable and the right hand side contains the factor variables of interest. Please use one-way layouts for the computation of the pseudo-ranks only. In case of higher-way layouts, please use a 'help factor' that#' shrinks the layout to a one-way design.
data	A data.frame, list or environment containing the variables in formula. The default option is NULL.
psranks	A header specifying the name of the pseudo ranks in the output data set.

Details

The pseudo-ranks are exported within a new column attached to the given data set.

References

Konietschke, F., Hothorn, L. A., & Brunner, E. (2012). Rank-based multiple test procedures and simultaneous confidence intervals. Electronic Journal of Statistics, 6, 738-759.

Brunner, E., Bathke, A. C., Konietschke, F. (2018). Rank and pseudo-rank procedures for independent observations in factorial designs. Springer International Publishing.

Happ, M., Zimmermann, G., Brunner, E., Bathke, A. C. (2020). Pseudo-ranks: How to calculate them efficiently in R. Journal of Statistical Software, 95(1), 1-22.

See Also

[rankFD](#)

Examples

```
data(Muco)
Muco2 <- psr(HalfTime~Disease,data=Muco, psranks="Mypseudos")
```

---

rank.two.samples   *A function for analyzing two-sample problems*

Description

The rank.two.samples() function calculates purely nonparametric rank-based methods for the analysis of two independent samples. Specifically, it implements the Brunner-Munzel test and its generalizations for the Nonparametric Behrens-Fisher Problem, that is, testing whether the relative effect p=P(X<Y)+1/2*P(X=Y) of the two independent samples X and Y is equal to 1/2. Range preserving confidence intervals (and corresponding test statistics) are available using Logit or Probit transformations. The function also implements studentized permutation tests and permutation based confidence intervals for p using any of the method above (see the details below). Furthermore, the Wilcoxon-Mann-Whitney test (exact and asymptotic) can be used to test the equality of the two distribution functions of the two samples. The user can specify whether confidence intervals for shift effects shall be computed. The rank.two.samples() function implements one-sided and two-sided tests and confidence intervals. You can plot the confidence intervals (for the relative effects) with the plot() function.

Usage

```
rank.two.samples(
  formula,
  data,
  conf.level = 0.95,
  alternative = c("two.sided", "less", "greater"),
  rounds = 4,
  method = c("t.app", "logit", "probit", "normal"),
  permu = TRUE,
  info = TRUE,
  wilcoxon = c("asymptotic", "exact"),
  shift.int = TRUE,
  nperm = 10000
)
```

Arguments

formula	A model [formula](#) object. The left hand side contains the response variable and the right hand side contains the factor variables of interest.
data	A data.frame, list or environment containing the variables in formula. The default option is NULL.
conf.level	A number specifying the confidence level; the default is 0.95.
alternative	A character string specifying the alternative hypothesis. One of "two.sided", "less", "greater". You can specify just the initial letter.
rounds	Value specifying the number of digits the results are rounded to. Default is 4 decimals.

method	specifying the method used for calculation of the confidence intervals. One of "t.app", "logit", "probit" or "normal".
permu	A logical variable indicating whether you want to compute a studentized permutation test.
info	A logical variable. Here, info = FALSE suppresses the output of additional information concerning e.g. the interpretation of the test results.
wilcoxon	asymptotic or exact calculation of Wilcoxon test.
shift.int	Logical, indicating whether or not shift effects should be considered.
nperm	Number of permutations used, default is 10000.

Details

The `rank.two.samples()` function calculates both transformed (logit or probit) and untransformed statistics (normal or t.app) for testing the null hypothesis p=1/2. If a studentized permutation test is performed, then the permutation distribution of the respective statistics are computed, see Pauly et al.(2016) for details. In any case, the function reports the point estimator and its estimated standard error, value of the test statistic, confidence interval and p-value. In case of separated samples, point estimator and standard error would be 0 and thus, test statistics would not be defined. In such a case, point estimator and its standard error are replaced by the numbers one would obtain if samples overlapped in a single point. A plot of the confidence interval can be obtained with the plot function.

Author(s)

Frank Konietschke Brunner, E., Bathke, A. C., Konietschke, F. (2018). Rank and Pseudo-Rank Procedures for Independent Observations in Factorial Designs. Springer International Publishing.

References

Brunner, E. and Munzel, U. (2000). The nonparametric Behrens-Fisher problem: Asymptotic theory and a small-sample approximation. Biometrical Journal 1, 17 - 21.

Kaufmann, J., Werner, C., and Brunner, E. (2005). Nonparametric methods for analysing the accuracy of diagnostic tests with multiple readers. Statistical Methods in Medical Research 14, 129 - 146

Pauly, M., Asendorf, T., Konietschke, F. (2016). Permutation-based inference for the AUC: a unified approach for continuous and discontinuous data.##' Biometrical Journal, 58(6), 1319 – 1337.

See Also

rankFD

Examples

```
data(Muco)
Muco2 <- subset(Muco, Disease != "OAD")
Muco2$Disease <- droplevels(Muco2$Disease)

twosample <- rank.two.samples(HalfTime ~ Disease, data = Muco2,
```

```
wilcoxon = "exact", permu = TRUE, shift.int = TRUE, nperm = 1000)
twosample <- rank.two.samples(HalfTime ~ Disease, data = Muco2,
   alternative = "greater", method = "probit", wilcoxon = "exact", permu = TRUE,
   shift.int = FALSE, nperm = 1000)
plot(twosample)
```

---

rankFD   *Rank-based tests for general factorial designs*

Description

The function implements purely nonparametric rank-based methods for the analysis of general
factorial designs. You can chose to use either classical ranks (mid-ranks) (effect="weighted")
or pseudo-ranks (effect="unweighted") for making inference. Pseudo-ranks are used by de-
fault. The package implements point estimators of relative effects (weighted and unweighted)
as well as test procedures (Wald-Type and ANOVA-Type statistics) for testing global null hy-
potheses formulated in either (i) distribution functions hypothesis="H0F" or (ii) relative effects
hypothesis="H0p". In case of one-way factorial designs, the function additionally computes the
Kruskal-Wallis test either with ranks or pseudo-ranks. In addition, multiple contrast tests (and si-
multaneous confidence intervals) for the main or interaction effects are implemented within the
contrast statement. You can either choose from pre-defined contrasts (options see below) or you
can provide your own user-defined contrast matrix. Both the Fisher-transformation (sci.method="fisher")
as well as a multivariate t-approximation (sci.method="multi.t") are implemented. The Fisher
approximation is used by default. To visualize the results, you can plot the simultaneous confi-
dence intervals using the plot.sci function. Furthermore, confidence interval plots for the main
or interaction relative effects (not simultaneous) are available within the plot function.

Usage

```
rankFD(
  formula,
  data,
  alpha = 0.05,
  CI.method = c("logit", "normal"),
  effect = c("unweighted", "weighted"),
  hypothesis = c("H0F", "H0p"),
  Factor.Information = FALSE,
  contrast = NULL,
  sci.method = c("fisher", "multi.t"),
  info = TRUE,
  covariance = FALSE,
  rounds = 4
)
```

Arguments

formula	A model [formula](#) object. The left hand side contains the response variable and the right hand side contains the factor variables of interest. An interaction term must be specified.
data	A data.frame, list or environment containing the variables in formula. The default option is NULL.
alpha	A number specifying the significance level; the default is 0.05.
CI.method	Either "logit" or "normal", specifying the method used for calculation of the confidence intervals.
effect	In case of weighted, then weighted (by sample sizes) relative effects are estimated using classical ranks (mid-ranks) of the data. Otherwise, in case of effect="unweighted", unweighted relative effects are estimated with pseudo-ranks. The default option is "unweighted" resulting in pseudo-rank statistics.
hypothesis	The null hypothesis to be tested, either "H0F" or "H0p". The option "H0F" computes tests for testing hypotheses formulated in terms of distribution functions. Otherwise, hypotheses in relative effects are tested. The latter allows for variance heteroscedasticity even under the null hypothesis of no treatment effect and thus covers the Nonparametric Behrens-Fisher problem.
Factor.Information	
	Logical. If TRUE, descriptive statistics with point estimators, standard error as well as confidence intervals for each main and interaction effect in the model are printed. The results can furthermore be plotted with the plot function.
contrast	a list containing the name of the main or interaction effect (written as group1:group2), a pre-defined contrast ("Dunnett", "Tukey", "Sequen", "AVE", "Changepoint", "Williams", "Marcus", "McDermott", "UmbrellaWilliams", "GrandMean") or a user-defined contrast matrix. If the contrast coefficients do not sum up to 0, or if their sum of absolutes differs from 2, the coefficients are normalized.
sci.method	Either "fisher" or "multi.t" as approximation method for the multiple contrast tests and simultaneous confidence intervals. The default option is "fisher".
info	Logical. If TRUE, additional output information and explanation is printed to the console.
covariance	Logical. If TRUE, the estimated covariance matrix of the vector of relative effects is computed.
rounds	Number of decimals of the output values. The default option is rounds=4 (4 decimals).

Details

The rankFD() function calculates the Wald-type statistic (WTS), ANOVA-type statistic (ATS) as well as multiple contrast tests for general factorial designs for testing the null hypotheses $H_0^F : CF = 0$ or $H_0^p : Cp = 0$. Almost every method explained in the comprehensive textbook from Brunner et al. (2019) is implemented in rankFD. The test procedures for testing null hypotheses in distribution functions have initially been proposed by Akritas et al. (1997), whereas methods for testing null hypotheses formulated in relative effects have been proposed by Brunner et al. (2017). We note that the multiple contrast test procedure using Fisher approximation computes critical and

p-values from a multivariate t-distribution with respective degrees of freedom. Simulation studies by Konietschke et al. (2012) demonstrated an accurate control of the type-1 error rate and the procedure is therefore recommended.

Value

A `rankFD` object containing the following components:

`Call`	Given response and factor names (formula)
`Descriptive`	Descriptive statistics of the data for all factor level combinations. Displayed are the number of individuals per factor level combination (size), the relative effect (Rel.Effect), Standard Error and 100*(1-alpha)% confidence intervals.
`WTS`	The value of the WTS along with degrees of freedom of the central chi-square distribution and p-value.
`ATS`	The value of the ATS, degrees of freedom of the central F distribution and the corresponding p-value.
`Kruskal-Wallis-Test`	The value of the Kruskal-Wallis test along with degrees of freedom and p-value. If effect="unweighted", the Kruskal-Wallis test using pseudo-ranks is computed. Otherwise, if effect="unweighted", the "established" Kruskal-Wallis test based on ranks is returned.
`MCTP`	Contrast matrix, local Results in terms of point estimates, standard error, value of the test statistic, (1-alpha)100 intervals as well as adjusted p-values. As a summary, the function also returns the global test decision by printing the maximum test statistic (in absolute value) as well as the (1-alpha) critical value from the multivariate T-distribution.
`Covariance.Matrix`	The estimated covariance matrix of the vector of the estimated relative effects. Note that the vector is multiplied by root N.
`Factor.Information`	Descriptive tables containing the point estimators, standard errors as well as (1-alpha)100 and interaction effects in the model. The confidence intervals are not simultaneous and for data descriptive purpose only.

References

Brunner, E., Bathke, A.C., Konietschke, F. Rank and Pseudo-Rank Procedures for Independent Observations in Factorial Designs. Springer International Publishing, 2018.

Brunner, E., Konietschke, F., Pauly, M., Puri, M. L. (2017). Rank-based procedures in factorial designs: Hypotheses about non-parametric treatment effects. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 79(5), 1463-1485.

Akritas, M. G., Arnold, S. F., and Brunner, E. (1997). Nonparametric hypotheses and rank statistics for unbalanced factorial designs. Journal of the American Statistical Association 92, 258-265.

Brunner, E., Dette, H., and Munk, A. (1997). Box-Type Approximations in Nonparametric Factorial Designs. Journal of the American Statistical Association 92, 1494-1502.

Konietschke, F., Hothorn, L. A., Brunner, E. (2012). Rank-based multiple test procedures and simultaneous confidence intervals. Electronic Journal of Statistics, 6, 738-759.

Examples

```
data(Coal)
model <- rankFD(Acidity ~ NaOH * Type, data = Coal, CI.method = "normal",
effect = "unweighted", hypothesis = "H0F")

data(Muco)
model.oneway <- rankFD(HalfTime ~ Disease, data = Muco, CI.method = "logit",
effect = "weighted", hypothesis = "H0p")
plot(model.oneway)
```

---

steel   *Steel-type multiple contrast tests*

Description

The function implements purely nonparametric Steel-type multiple contrast tests for either making many-to-one (Dunnett-type) or all pairwise (Tukey-type) comparisons. Null hypotheses are formulated in terms of the distribution functions.

Usage

```
steel(
  formula,
  data,
  control = NULL,
  alternative = c("two.sided", "less", "greater"),
  info = TRUE,
  correlation = TRUE
)
```

Arguments

formula	A model [formula](#) object. The left hand side contains the response variable and the right hand side contains the factor variable of interest.
data	A data.frame, list or environment containing the variables in formula. The default option is NULL.
control	Specification of the control group for making many-to-one-comparisons. If NULL, all-pairwise comparisons are performed.
alternative	Specification of the direction of the alternative. Default is two-sided.
info	Logical. If TRUE, additional output information and explanation is printed to the console.
correlation	Logical. If TRUE, the correlation matrix is printed.

Details

The steel() function calculates the Steel-type tests as explained by Munzel, U., Hothorn, L. A. (2001). A unified approach to simultaneous rank test procedures in the unbalanced one-way layout. Biometrical Journal: Journal of Mathematical Methods in Biosciences, 43(5), 553-569.

Value

A list containing the following components:

Data.Info	Groups and sample sizes of the data
Analysis	Data frame containing the test results (comparison, relative effect estimator, standard error, test statistic and p-value.)
Correlation	Estimated correlation matrix

References

Brunner, E., Bathke, A.C., Konietschke, F. Rank and Pseudo-Rank Procedures for Independent Observations in Factorial Designs. Springer International Publishing, 2018.

Munzel, U., Hothorn, L. A. (2001). A unified approach to simultaneous rank test procedures in the unbalanced one-way layout. Biometrical Journal: Journal of Mathematical Methods in Biosciences, 43(5), 553-569.

Konietschke, F., Hothorn, L. A., Brunner, E. (2012). Rank-based multiple test procedures and simultaneous confidence intervals. Electronic Journal of Statistics, 6, 738-759.

Examples

```
data(Muco)
model.oneway <- steel(HalfTime ~ Disease, data = Muco,info=TRUE,correlation=TRUE)
```

WMWSSP	*Sample size computation methods for the Wilcoxon-Mann-Whitney test based on pilot data*

Description

The function implements the sample size formula proposed by Happ et al. (see reference below). It estimates the sample size needed to detect the effect with pre-defined power at significance level alpha based on pilot data.

Usage

```
WMWSSP(x1, x2, alpha = 0.05, power = 0.8, t = 1/2)
```

Arguments

x1	advance information for the first group
x2	advance information for the second group
alpha	two sided type I error rate
power	power with the sample sizes of each group
t	proportion of subjects in the first group

Value

Returns a data frame

References

Brunner, E., Bathke A. C. and Konietschke, F. Rank- and Pseudo-Rank Procedures in Factorial Designs - Using R and SAS. Springer Verlag. Happ, M., Bathke, A. C., & Brunner, E. (2019). Optimal sample size planning for the Wilcoxon-Mann-Whitney test. Statistics in medicine, 38(3), 363-375.

Examples

```
x1 = c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 2)
x2 = c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 3)
WMWSSP(x1,x2,0.05,0.8,0.5)
```

Index